The Importance of Sample Attrition in Life Cycle Labor Supply

Ziliak, J.P.; Kniesner, T.J.

Publication date:
1996

Citation for published version (APA):
Ziliak, J. P., & Kniesner, T. J. (1996). *The Importance of Sample Attrition in Life Cycle Labor Supply*. (CentER Discussion Paper; Vol. 1996-46). Vakgroep CentER.

The Importance of Sample Attrition in Life Cycle Labor Supply Estimation

James P. Ziliak
Assistant Professor of Economics
University of Oregon, Eugene

and

Thomas J. Kniesner*
Professor of Economics
Indiana University, Bloomington

June 5, 1996

Abstract: We examine the importance of possible non-random attrition to an econometric model of life cycle labor supply including joint nonlinear taxation of wage and interest incomes and latent heterogeneity. We use a Wald test comparing attriters to nonattriters and variable addition testing based on formal models of attrition. Results from the Panel Study of Income Dynamics are that non-random panel attrition is of little concern for prime-aged male labor supply estimation because the effect of attrition is absorbed into the fixed effects. Attrition is less econometrically influential than research design decisions typically taken for granted; the wage measure or instrument set has a much greater impact on the estimated labor supply function of prime-aged men than how one includes panel attrition.

Key words: attrition bias, Panel Study of Income Dynamics, life cycle labor supply, nonlinear income tax, generalized method of moments

* We thank three anonymous referees, Joe Altonji, Greg Duncan, Arthur Goldberger, Mark Gritz, Joel Horowitz, Arie Kapteyn, Tom MaCurdy, Randy Olsen, Peter Schmidt, Pravin Trivedi, Marno Verbeek, Jeff Wooldridge, and participants at the BLS Conference on Attrition Bias in Longitudinal Surveys, the Fifth International Conference on Panel Data, the Annual Midwest Econometrics Group Meeting, and the Tilburg University CentER-VSB Saving Workshop for constructive criticisms. Our gratitude for financial support goes to Indiana University's Office of International Programs and Office of the Vice President for Research and the University Graduate School.

Communications: Thomas J. Kniesner
Department of Economics, Ballantine Hall (901)
Indiana University
Bloomington, IN 47405-6620 USA
Fax: (812) 855–3736 Phone: (812) 855–1021
Internet: Kniesne@Ophelia.UCS.Indiana.Edu

1. Introduction

Discontinuous participation in a panel survey, known as attrition, can happen for several reasons. Some people move and cannot be traced, others become institutionalized or die, and others are rotated out by a sampling design. Because attrition is cumulative it becomes a potentially more serious econometric concern as a panel continues. As a point of reference, 40 percent of the original 1968 Panel Study of Income Dynamics (PSID) sample had left the panel by 1981 (Becketti et al. 1988). Our research examines the importance of how (whether) one accounts for panel attrition when estimating the life cycle labor supply of prime-aged men with data from the PSID.

The key issue is whether attrition is (non)random. Random attrition can happen because the panel rotates new participants into the sample on a regular basis, as in the Survey of Income and Program Participation (SIPP). If attrition is random then parameters estimated from a panel of nonattriters are consistent, although there may be efficiency gains by including the incomplete information from the attriters (Hsiao 1986). If attrition is not random but systematically related to the model's endogenous variables then econometric estimates based on nonattriters only are inconsistent. In the terminology of the statistics literature how one deals with attrition in the estimation hinges on whether attrition is random versus non-ignorable or informative (Diggle and Kenward 1994).

In their widely cited study Hausman and Wise (1979) examined the effects of non-random attrition on earnings equations estimated from the Gary Income Maintenance Experiment. Because high-wage experimentals received no treatment related income Hausman and Wise argued that the decision by high-wage experimentals to leave the experiment could be related to latent heterogeneity that made them naturally high-wage earners. They emphasized that ignoring the relation between the decision to attrite and latent heterogeneity could lead to inconsistent estimates of the earnings equation in general and of the treatment effect in particular.[1]

There is a similarly compelling reason for studying the importance of attrition to life cycle labor supply estimates in the presence of income taxation. Our specification of the worker's labor supply schedule includes the net (after-tax) wage plus beginning-of-period and end-of-period net wealth as regressors, all of which are endogenous to labor supply. If the decision to attrite comes from unobserved preferences to work (earn income) then labor supply parameters and subsequent deadweight loss calculations are inconsistently estimated if attrition is not part of the econometric model’s structure.

Economists know little about the importance of attrition to labor supply estimates. When estimating a static employment status model with the Seattle and Denver negative income tax experiments’ data controlling for possible endogenous sample composition made no significant difference to the estimated treatment effect (Robins and West 1986). Our research has little in common with Robins and West because we estimate a life cycle consistent labor supply model with nonlinear income taxes and latent worker heterogeneity. In the research most similar to ours Zabel (1994) found significant selection correction terms in the labor force participation of white men, but structural labor supply parameters that did not change significantly after correcting for attrition. What distinguishes our research from Zabel's is that we include joint nonlinear taxation of wage and nonwage income, examine various specifications of the attrition (panel continuation) probability equation in the context of a two-step GMM estimator, and offer a wider context in which to judge the econometric importance of how one adds sample attrition to an econometric labor supply model.

To elaborate, we use a sequential econometric procedure to infer confounding effects of non-random attrition in a model of the labor supply of prime-aged men. First, we study

[1] Their focal econometric result (the estimated NIT treatment effect) did not change after careful modeling of attrition, however.

the econometric importance of attrition in a way that does not require specifying an attrition process through a simultaneous equations Wald test for structural change (Lo and Newey 1985). The Wald test indicates whether persons who attrite are different than persons who do not attrite in terms of labor supply. Next we examine attrition's importance to labor supply estimation through an explicit model of the attrition decision within a two-step estimator.[2] The two-step procedure first estimates an equation for the probability of panel continuation then adds a transformation of the continuation probability as an additional regressor in the labor supply equation. Evidence of informative attrition includes significance of the coefficient of the transformed survival probability regressor in the second-stage labor supply regression.

Summary. We find that the economic parameters of interest, the estimated net wage and wealth effects on labor supply, are generally unaffected by how (whether) the researcher adds attrition to the econometric model because attrition is adequately modeled as a fixed effect. As a point of reference we further demonstrate that labor supply estimates for prime-aged men are more sensitive to other decisions researchers make that are frequently taken for granted, such as choosing the wage rate measure or instrument set.

2. Econometric Background

We begin by describing the economic model underlying our structural econometric model of labor supply while for the moment maintaining maximum econometric generality when describing panel attrition. If both capital and wage incomes are taxed nonlinearly the associated intertemporally nonseparable lifetime budget constraint can be included by conditioning labor supply on the worker's asset positions at the beginning and end of each period (Blomquist 1985). The econometric model we use to study attrition's

[2]A two-step method is preferred to joint maximum likelihood, which forces a positive estimated net wage coefficient and a negative estimated wealth coefficient rather than permitting the researcher to verify that the behavior implied by economic theory is appears in the data (MaCurdy, Green, and Paarsch 1990).

consequences is a linear labor supply function that conditions on current and lagged assets, which is a life cycle consistent model under two stage budgeting.[3]

2.1 Life Cycle Consistent Labor Supply

To make our exercise maximally informative we anchor the underlying economics to the seminal research on labor supply and nonlinear taxes (Hausman 1981). We estimate a linear labor supply function allowing nonstochastic latent worker heterogeneity

$$h_{it} = \alpha\omega_{it} + \phi A_{it} + \delta A_{it-1} + \gamma' X_{it} + \eta_i + \xi_{it}\,, \qquad \textbf{(1)}$$

where i indexes workers, t indexes time period, h is annual hours worked, ω is the net after-tax hourly wage, A is net wealth, X is a vector of time-varying demographics affecting intratemporal preferences for work, $\Lambda = [\alpha, \delta, \phi, \gamma']$ are the parameters of intratemporal preferences, and the error term ξ_{it} is iid with mean zero and constant variance. The net wage and assets are endogenous because the marginal tax rate depends on contemporaneous hours worked through earnings. Finally, the time-invariant worker-specific effect, η_i, is generally not independent of the regressors because life cycle wealth has person-specific components unknown to the econometrician.[4]

2.2 Incorporating Nonlinear Income Taxes

Although the most influential econometric research on the labor supply effects of income taxes has applied the maximum likelihood approach to represent the piecewise linear budget constraint (Hausman 1981), maximum likelihood rests on empirically unsupported assumptions. Maximum likelihood requires that the gross wage and gross wealth be exogenous to labor supply. To ensure positive probabilities and a well-behaved likelihood function, maximum likelihood also regulates the allowable set of labor supply responses, which forces a nonnegative estimated wage effect and a nonpositive estimated

---

[3]Our estimating equation identifies intratemporal, but not intertemporal, preferences. For a two-step estimator that also recovers intertemporal preference parameters see Ziliak and Kniesner (1995).

[4]We limit latent worker-specific labor supply heterogeneity to the intercept. Allowing worker heterogeneity in the coefficients of endogenous wages or wealth given the complexity of nonlinear income taxation and possible endogenous attrition is best left for the future. See Kniesner and Li (1996) for a general econometric model of labor supply with heterogeneous response parameters.

asset effect, which need not be the case in a life-cycle labor supply model with time nonseparabilities caused by nonlinear income taxes as we consider here (Blomquist 1985; MaCurdy, Green, and Paarsch 1990). Because of the econometric complexity and the stringent and possibly incorrect ex ante restrictions that maximum likelihood places on estimated labor supply parameters an instrumental variables type estimator such as GMM that we use is most preferred.[5]

Reported taxable income is relatively free of measurement error in the typical micro data set such as the PSID so that the marginal tax rate can be tracked closely by a differentiable polynomial in taxable income (MaCurdy, Green, and Paarsch 1990). A differentiable marginal rate can also be integrated to infer total taxes needed to construct net wealth. Adopting the differentiable marginal tax rate approach of MaCurdy, Green, and Paarsch in constructing net wages and assets also simplifies parameterizing the limited tax base for social security taxes. During our sample period most states also had progressive income tax schedules where about 75 percent of the states used federal Adjusted Gross Income or federal taxable income as their bases. We judge the possible labor supply effects of state income taxes too important to ignore but too complicated to include completely. In our labor supply estimates we augment the worker's federal marginal tax and social security tax rates with an average state tax rate that is the ratio of individual state income tax collections to AGI in the state.[6]

2.3 Incorporating Sample Attrition

Hours worked by person *i* at time *t* in (1) can be written compactly as

$$h_{it} = \Lambda Z_{it} + \upsilon_{it}, \qquad \textbf{(2)}$$

where $Z_{it} = [\omega_{it}, A_{it-1}, A_{it}, X_{it}]'$, $\Lambda = [\alpha, \delta, \phi, \gamma']$, and $\upsilon_{it} = \eta_i + \xi_{it}$, collects latent heterogeneity (η_i) and the overall random shock (ξ_{it}). When panel non-response is

[5]GMM also requires only information on the effective marginal tax rate, which substantially eases computation.
[6]For more discussion see Ziliak and Kniesner (1995).

possible (2) is observed when an indicator function, $r_{it} = 1$. That is, we now admit the possibility of attrition ($r_{it} = 0$) such that only nonattriters are observed for all t in the PSID.

The indicator function for sample survival obeys the relationship with a latent variable $r_{it}{*}$ such that

$$r_{it} = 1 \text{ if } r_{it}^{*} > 0;\ r_{it} = 0 \text{ otherwise,} \quad \textbf{(3)}$$

where

$$r_{it}^{*} = \beta Z_{it}^{*} + \upsilon_{it}^{*}. \quad \textbf{(4)}$$

The elements of Z_{it}^{*} are regressors that explain the outcome of continuing in the panel, some of which also may influence labor supply, and υ_{it}^{*} is the error term. We discuss the details of the panel continuation equation shortly.

Two-Step Estimation. A common econometric approach for handling endogenous worker heterogeneity is as a nonstochastic (fixed) effect. One way to estimate a fixed effects model is to use the within estimator, and another way is to estimate the model in first-differences. A labor supply equation conditioned on the net wage and current and lagged assets as in (1) causes a within estimator with predetermined instruments to be inconsistent. Because the tax rate depends on current hours worked, deviations from the individual time-series means needed as regressors in the within estimator will not be independent of the overall labor supply errors, ξ_{it}. The first-differences estimator we use for labor supply (1) is consistent (Keane and Runkle 1992).[7]

Estimating a structural model in the presence of non-randomly changing sample composition without controls for a possibly endogenous panel continuation process yields inconsistent parameter estimates (Heckman 1979). Heckman suggested a consistent two-step estimator where the first step produces a sample composition equation, and the estimated sample composition parameters are then used to construct an additional

[7]Another problem with the within estimator is the difficulty of finding good instruments whereas in the first differences estimator the endogenous variable lagged two or more periods can be used as instruments. On the downside first-differencing may exacerbate any measurement error (Altonji 1986). If our instruments are uncorrelated with the measurement error the parameter estimates are still consistent.

regressor for the second step regression recovering the structural parameters of interest from a sample of complete observations. Heckman's two-step method is readily extended to panel data with possible non-random attrition.[8]

Attrition from the PSID is an absorbing state; once someone leaves the panel they are gone for good. The implication of once-and-for-all attrition is that panel continuation cannot be viewed simply as a continuous binary outcome. We must instead treat attrition as a discrete hazard process. During each period every observation comprises the risk set, where risk is the probability of continued participation in the panel. As soon as attrition takes place a worker is no longer part of the risk set. The dependent variable in our discrete hazard function equals one for each period someone is in the sample and equals zero the first (and only) time a worker departs the PSID.

We use the first-difference form to eliminate the person-specific latent heterogeneity. The second-step structural supply equation we estimate that corrects for the likelihood of continued panel participation under two possible alternative (normal versus logistic) probability processes is

$$\Delta h_{it} = \alpha \Delta w_{it} + \phi \Delta A_{it} + \delta \Delta A_{it-1} + \gamma' \Delta X_{it} + \sigma \hat{\lambda}_{it} + \Delta \varepsilon_{it} \qquad \textbf{(5)}$$

with
$$\hat{\lambda}_{it} = \begin{cases} \phi(\hat{\beta}' Z^*_{it}) / \Phi(\hat{\beta}' Z^*_{it}) \\ or \\ -[\ln(\Lambda(\hat{\beta}' Z^*_{it})) - \hat{\beta}' Z^*_{it}(1 - \Lambda(\hat{\beta}' Z^*_{it}))] / \Lambda(\hat{\beta}' Z^*_{it}) \end{cases} \qquad \textbf{(6)}$$

where $\phi(\bullet)$ is the pdf of the normal distribution, $\Phi(\bullet)$ is the cdf of the normal distribution, and $\Lambda(\bullet) = \exp(\hat{\beta}' Z^*_{it}) / [1 + \exp(\hat{\beta}' Z^*_{it})]$ is the cdf of the logistic distribution (Heckman and MaCurdy 1986).

The GMM Estimator. Define the function $g(Z, D; \Lambda)$ as

$$g(Z, D; \hat{\Lambda}) = D'(h - Z\hat{\Lambda}) \equiv D'\xi\,, \qquad \textbf{(7)}$$

---

[8]One does not always need a parametric form for the attrition process and can consider nonparametric alternatives (Manski 1989, 1993).

where Z is the ($N(T-1)\times L$) matrix of regressors in the labor supply function (1), D is an ($N(T-1)\times K$) matrix of instruments with $K \geq L$, h is the ($N(T-1)\times 1$) vector of hours worked , and $\hat{\Lambda}$ is the vector of ($L\times 1$) preferences parameters that are the coefficients in the linear labor supply function, $[\alpha,\delta,\phi,\gamma']$.

The criterion function we minimize in our GMM first-differences model is

$$J_T = g(Z,D;\Lambda)' S_{gg}^{-1} g(Z,D;\Lambda)\,, \tag{8}$$

where S_{gg} is an optimal weighting matrix, $D'E(\xi\xi')D$. Initial consistent estimates for the vector error ξ come from a consistent but suboptimal weighting matrix, the identity matrix. Solving the criterion function for the feasible GMM estimator gives

$$\hat{\Lambda} = [Z'D\hat{S}_{gg}^{-1}D'Z]^{-1} Z'D\hat{S}_{gg}^{-1}D'h, \tag{9}$$

which has the estimated covariance matrix for large N and finite T

$$Var(\hat{\Lambda}) = [\mathrm{Z}'\mathrm{D}\hat{\mathrm{S}}_{\mathrm{gg}}^{-1}\mathrm{D}'\mathrm{Z}]^{-1}. \tag{10}$$

Estimating the first-differenced labor supply (5) as a way of coping with latent heterogeneity and possible life cycle rational expectations creates an MA(1) process in the transformed random disturbance, $\xi_t - \xi_{t-1}$, which influences the functional form of the weighting matrix, S_{gg} (Maeshiro and Vali 1988). The weighting matrix in our GMM first-differences model $\hat{S}_{gg}$ is the sum of a conditionally heteroskedastic matrix ($\hat{\Omega}_0$) and an autocorrelation matrix ($\hat{\Omega}_1$) such that

$$\hat{S}_{gg} = \hat{\Omega}_0 + [\hat{\Omega}_1 + \hat{\Omega}_1'], \tag{11}$$

where
$$\hat{\Omega}_0 = (1/\,N(T-1))\sum_i \sum_t (D_{it}'\hat{\xi}_{it}\hat{\xi}_{it}' D_{it})\,, \tag{12}$$

$$\hat{\Omega}_1 = (1/\,N(T-1))\sum_i \sum_t (D_{it}'\hat{\xi}_{it}\hat{\xi}_{it-1}' D_{it-1})\,, \tag{13}$$

$i = 1,\ldots,N$, and $t = 1,\ldots,T$.[9] Information dated *t–2* and earlier can be instruments in light of the MA(1) errors in the first-differenced life cycle consistent labor supply (Griliches and Hausman 1986). The first differencing, lagged instruments, and correcting for the MA(1) term in the weight matrix together mean we can only use observations present in four

---

[9]When the weighting matrix is not positive definite we use a method of modified Bartlett weights (Newey and West 1987b).

consecutive waves in estimation so that we actually use N(T–3) observations in the estimation of the labor supply parameters and the covariance matrix elements.

A basic specification test in our GMM estimator is a test of validity of the overidentifying restrictions. The overidentification test statistic is the value of the criterion function, J_T, at the final GMM parameter estimates and is distributed as $\chi^2(p)$, where p is the number of instruments less regressors. In general, restrictions can be tested with the objective function based test of the form

$$J = T[J_T(\hat{\Lambda}_r) - J_T(\hat{\Lambda})] \sim \chi^2(p), \qquad \textbf{(14)}$$

where the subscript r indicates the restricted model, and the p degrees of freedom in the computed chi-squared statistic is the number of restrictions imposed (Newey and West 1987a).

2.4 Examining the Econometric Significance of Attrition

As we have seen a complete model of labor supply and sample attrition can be computationally cumbersome. Tests of whether there seems reason for econometric concern over attrition from the panel are useful because they can indicate if it is even necessary to model the attrition process itself (Verbeek and Nijman 1992). A Wald test for non-random attrition can be a useful starting point for models where attrition bias is of concern.

A Wald Test. A sufficient condition for ignorable or non-informative attrition in the fixed effects labor supply model estimated in first-differences is $E[\xi_{it}^d | r_{it}, r_{it-1}] = 0$, where the superscript d indicates first-differences (Verbeek and Nijman 1992). Even though attrition may have an individual effect common to labor supply, η_i, ignoring attrition will not introduce selectivity bias in the fixed-effects estimator when attrition is independent of ξ_{it}^d. An attrition effect in labor supply that is time invariant is captured in the fixed effect and swept out by first-differencing.

The Wald procedure for a linear simultaneous equations system tests whether the underlying labor supply process is the same for workers who attrite as for workers who continue in the panel survey (Lo and Newey 1985). If $V(A)$ is the estimated covariance matrix for attriters and $V(NA)$ is the estimated covariance matrix for non-attriters then the Wald test statistic to use is

$$W = (\hat{\Lambda}(A) - \hat{\Lambda}(NA))'[V(A) + V(NA)]^{-1}(\hat{\Lambda}(A) - \hat{\Lambda}(NA)) \sim \chi^2(k), \quad \textbf{(15)}$$

where k is the number of regressors in the first-differenced labor supply equation.

Variable Addition Test. Variable addition as model specification testing has gained widespread acceptance (Davidson and MacKinnon 1993). The crux of variable addition tests is that under the null hypothesis the added variable(s) are exogenous to the structural equation disturbance term.[10] A test of the null hypothesis that attrition is ignorable that we use is a *t*-test of significance of $\hat{\sigma}$ in (2).

3. Data

We use data from Waves I–XXII (interview years 1968–1989) of the Panel Study of Income Dynamics to estimate labor supply parameters and examine the econometric consequences of panel attrition. The PSID began in 1968 with about 4800 households and over 18,000 persons; by the 1989 wave the PSID had over 7000 families and 37,000 persons. About 61 percent of the initial PSID households were a random sample of the U.S. population selected by the Survey Research Center (SRC), and the remaining 39 percent of the initial PSID households were a sample of the low-income families drawn from the Survey of Economic Opportunity (SEO). Because the SEO oversampled the poor, researchers pooling the SRC and SEO samples should weight the first and second moments of population statistics. There is much disagreement on the merits of weighting a regression model, and in a sample of both attriters and nonattriters it is even unclear which weight to use for the population statistics (Hoem 1989). On the one hand it seems reasonable to use the weight from the most recent wave that a person contributes data (Hill 1992, p. 61). On the other hand it seems appropriate to use the original 1968

[10]This rules out attrition (panel continuation) as a function of wealth and lagged wages which, through wealth effects, are both endogenous to labor supplied.

weights, which were designed to adjust for stratified sampling (Lillard 1989, p. 508). We follow Lillard's suggestion and use the 1968 weights for the population statistics reported in Table 2 and do not weight the data for the econometric models of labor supply.[11]

3.1 Samples

We constructed two samples from the overall PSID: a balanced panel and an unbalanced panel. In the balanced panel there are data on all regression variables in every year that the person is a panel participant. In the unbalanced panel only a person year is absent when a missing value occurs. Although in a balanced design there is a substantial loss of observations the balanced design helps one avoid mingling the econometric importance of wave non-response with item non-response.

Our selection rules for the balanced panels are similar to other research: continuously working, non self-employed prime-aged men ages 25–43 in 1968. Because the oldest worker is no older than 64 we can safely ignore possible endogenous retirement decisions. We permit marital status to vary over the sample period and allow marital status change to be predetermined with labor supply (Johnson and Skinner 1986). In addition, we do not include non-sample members, persons who marry into the sample, or persons who attrite due to death because the data generating process may distort our tests of attrition's consequences (Lillard 1989). The selection criteria we used created (1) a balanced panel with 200 attriters contributing 711 persons years and 89 nonattriters contributing 1958 person years and (2) an unbalanced panel with 303 attriters contributing 1867 persons years and 315 nonattriters contributing 7100 person years.[12]

[11] For discussion of the PSID sample design, composition, attrition rates, and weighting see Becketti et al. (1988), Lillard (1989), and Hill (1992).

[12] We also relaxed the selection criterion that a man work positive hours in every year. After applying all other missing data screens allowing annual hours worked of zero increased sample size by only three percent so we ignored work status changes as a selection criterion issue of much research interest.

3.2 Key Variables

The variables in our econometric models are defined in Table 1. To compute real wages, income, interest rates, and assets we used the annual average for the year before the interview of the base 1987 GDP deflator for personal consumption expenditures. We now discuss the key labor supply regression variables that a labor supply researcher must construct when using the PSID, which are the wage rate, wealth, and taxes.

Wage Rate. We use multiple measures of the gross and net (post-tax) hourly wage rate: (1) average hourly earnings computed as the ratio of annual earnings to annual hours worked, (2) average hourly earnings computed as the ratio of annual earnings to annual weeks worked times usual hours worked per week, and (3) the hourly pay the respondent reports. It is well documented that average hourly earnings computed with the dependent variable of the labor supply regression induces a so-called negative division bias into the labor supply wage parameter (Conway and Kniesner 1994, Ziliak and Kniesner 1995). By using the three different wage measures we highlight the importance of an accurate wage measure compared to how one considers attrition in labor supply model estimation.

Wealth. Because the PSID does not have detailed information on either consumption or saving constructing the components of wealth is time consuming. We define wealth as the sum of liquid and illiquid assets. Liquid assets include nominal rent, interest, and dividend incomes capitalized by a nominal interest rate (Runkle 1991). We divided the first $200 of rental income by an annual average passbook savings rate and capitalized interest income exceeding $200 by the annual average 3-month T-Bill rate. Because the value of liquid assets understates the total wealth of a household we added an illiquid component of assets defined as the value of home equity. We measured home equity as the difference between house value and outstanding loan principal remaining.[13] The PSID

[13]Because principal remaining is missing for all persons in 1968, 1973–1975, and 1982 we follow the convention of the PSID staff and take 90 percent of the previous year's principal. Because data on home equity are still not available in the first year, 1968, we first set home equity in 1968 to its 1969 value then set home equity in 1968 to zero. Imputing 1968 illiquid wealth is less important than it may seem. The

collected comprehensive wealth data in 1984 and 1989, including data on home equity, net value of other real estate, net value of vehicles, net value of a farm or other business, and net value of other assets. The more direct measure of total wealth from the PSID has been used by others (Hubbard, Skinner, and Zeldes 1994, 1995). Variation in our measure of liquid wealth explains about half the variation in total wealth and including home equity makes the variation in our measure of wealth explain 80 of the variation in directly measured wealth (Ziliak 1994). The ability of our wealth measure to track total wealth when measured independently is our justification for including both liquid and illiquid wealth components in our definition of wealth.[14] Our wealth summary statistics are comparable to wealth measures from the Survey of Income Program Participation (Engen, Gale, and Scholz 1994).

Taxes. In constructing taxable income for each year we assumed that each person filed either a joint tax return if married or a head-of-household return if not married. Adjusted gross income (AGI) is the sum of the labor earnings for the man along with his interest income. Taxable income in each year is defined as adjusted gross income less deductions and exemptions.

The PSID records the number of exemptions (dependents) taken for tax purposes. For years before 1983 we followed the convention established in the PSID for computing deductions. Using information from the Internal Revenue Service's *Statistics of Income*, we generated the typical value of itemized deductions based on adjusted gross income. Using 1968 as an example of what we did in 1968–1983, if AGI was less than $5,000 in 1968 then the percent itemized from AGI was set to 23 percent; if AGI was greater than $5,000 but less than $10,000 then the percent itemized from AGI was set to 19 percent.

---

value of home equity in 1968 comes into the model only as an instrument in the MA(1) part of the error term. Because results were similar for the two 1968 asset imputations we tabulated only the results where home equity was set to zero in 1968.

[14]For completeness we run parallel regressions with assets defined first as the sum of illiquid and liquid assets and second as liquid assets alone.

We followed the process of imputing deductions based on national averages until AGI was greater than $20,000 when the average percent itemized from AGI is 15 percent. We imputed deductions similarly for the other years.

In the years 1968–1977 we constructed taxable income as follows. Using 1968 as an example, we first compared the standard deduction, which was 10 percent of AGI in 1968, to the so-called minimum standard deduction, which was $200 plus $100 times the number of exemptions in 1968. We then took the larger of total itemized deductions and the minimum standard deduction and compared it to percent itemized on average from AGI. If either the standard deduction or the minimum standard deduction were largest we then computed taxable income as AGI less exemptions and the greater of the standard deduction and minimum standard deduction. If the average percent itemized were largest then we computed taxable income as AGI less exemptions and percent itemized.

The values for the standard deduction, minimum standard deduction, and percent itemized varied over the years. Beginning in tax year 1978 until tax year 1987 the minimum standard deduction was eliminated, and the standard deduction was built into the tax tables. For 1978–1987 we took the difference between itemized deductions and the standard deduction, known as excess itemized deductions. If they were positive then we subtracted excess itemized deductions from adjusted gross income to compute taxable income; if excess itemized deductions were nonpositive then taxable income is simply AGI minus exemptions. Since the Tax Reform Act of 1986 (TRA86) the standard deduction is no longer built into the tax tables. For the years when TRA86 rules apply when excess itemized deductions were nonpositive we then computed taxable income as AGI less exemptions and the standard deduction.

3.3 Summary Statistics

Table 2 presents weighted and unweighted selected summary statistics for attriters and nonattriters in the balanced and unbalanced samples. Table 2 illustrates that, on average, the attriters are younger, work fewer hours, earn a lower hourly wage, have a lower marginal tax rate, have lower liquid and illiquid assets, have less education, are more likely to be black, less likely to own a home, less likely to be married, and are present for about 10 of the 22 possible waves of data.

4. Results

As described earlier we used a Wald test and a variable addition procedure to examine the econometric importance of possible non-random panel attrition in labor supply estimation. Although their coefficients are not tabulated in the interest of space, each labor supply specification includes as control variables the head's age, number of children in the home, health status, and marital status. The instrument set includes a constant, age, age^2, age$*$education, union status, health status, home ownership, marital status, and number of children at home, all dated t–1 and t–2, plus gross and net wage, net wealth, net virtual wealth, and the net 3-month T-Bill yield, all dated t–2. Based on results from Ziliak and Kniesner (1995) we also include time dummies in the instrument set, which makes a maximum number of 40 instruments.[15] For every labor supply function we present estimates using both the balanced and unbalanced panels from the unweighted joint SRC/SEO data in the PSID.

4.1 Wald Test Results

Tables 3 and 4 contain estimates of the life cycle consistent labor supply equation parameters separately for attriters and non-attriters. For brevity we report only the wage and wealth coefficients and the associated wage elasticities computed at the means of the sample used in estimation. Because wealth is endogenous the compensated wage elasticity is a first-order approximation to the true compensated wage elasticity (MaCurdy 1983).

[15]There are 37 instruments in the balanced attriter sample because the last year a worker may be present is 1986; there are 39 instruments in the unbalanced attriter sample because the last year a worker may be present is 1988.

All results in Table 3 use average hourly earnings as both the wage regressor and an instrument. The corresponding models in Table 4 also use average hourly earning as the wage regressor but instead use average earnings per usual hours worked in the instrument set. To compare further the robustness of the Wald test results we also study the importance of the choice of wealth measure. Estimates in the columns labeled (1) in Tables 3 and 4 are based on total wealth, which includes liquid assets plus home equity, and the results in the columns labeled (2) are based on liquid wealth, which includes only liquid assets.

Balanced Panel. In general, the *J*-statistic does not reject the null hypothesis that the overidentifying restrictions hold for the labor supply models estimated with the balanced panels. Visually comparing the estimated wage coefficients in Tables 3 and 4 suggests differing labor supply responses across attriters and non-attriters. Wald test results indicate no significant difference between attriters and non-attriters, however. Recall that there are only 89 non-attriters in the balanced panel. As noted earlier we need at least four years of data to estimate the model so that only about 100 attriters remain when estimating with the balanced panel. The substantial, yet statistically insignificant, difference between the estimated labor supply functions of attriters and non-attriters stems from the small sample sizes in the balanced panels.

The first four columns of Tables 3 and 4 also illustrate that the labor supply parameters and their standard errors are sensitive to the wage measure in the instrument set. Replacing average hourly earnings with average earnings per usual hours worked in the instrument set in Table 4 leads to a substantial relative efficiency loss in estimation for both attriters and non-attriters and does not solve the negative division bias problem in the estimated wage effect. Using Bartlett weights does not solve the problem of a negative definite variance-covariance matrix for the attriters in the second column of Table 4, which makes the Wald test statistic undefined. Likewise, removing home equity from the measure of wealth causes an undefined (negative) *J*-statistic for the non-attriters in the third column of Table 4. We emphasize that the results in the first four columns of Tables 3 and 4 are based on small samples, which is a fact of life when constructing a balanced panel of prime-aged men from the PSID. Although the balanced panel allows us to focus on wave non-response as separate from item non-response in the PSID the small sample sizes result in unrobust parameter estimates and low power test statistics for comparing the labor supply functions of attriters versus non-attriters.

Unbalanced Panel. Comparing Tables 3 and 4 illustrates the relative efficiency gain from moving to the unbalanced panel. The three to four times larger samples in the unbalanced panel produce more similar parameter estimates across columns and more powerful test statistics than in the balanced panel. In both Tables 3 and 4 non-attriters satisfy Slutsky integrability conditions but the attriters do not. The Wald statistics indicate a significant difference in the labor supply equations of attriters and non-attriters that is robust to the different wealth measures and wage instruments. The compensated wage estimates are theoretically correct for non-attriters but inconsistent with labor supply theory for attriters, which contrasts with the result that the overidentifying restrictions are rejected for non-attriters but not rejected for attriters in the unbalanced panel.

Summary. Although the unbalanced panel may muddy discussion of wave non-response and item non-response relative to the balanced panel the larger sample sizes in the unbalanced panel are necessary to have confidence in the estimated wage and asset parameters and overall J and Wald test statistics. We conclude that the data generating process for labor supply may be different for workers who left the PSID compared to workers who continued. To examine more closely the differences between attriters and non-attriters and their econometric consequences for estimating male labor supply we now move to two-step selection corrected labor supply models.

4.2 The Panel Continuation Process and Selectivity Corrected Results

Table 5 presents estimates of the discrete hazard functions for continued participation in the PSID. For completeness we estimate both discrete probit and logit hazard models with and without the number of waves completed as a regressor for both the balanced and unbalanced samples.

To estimate the discrete hazard functions we assemble the data in person years such that in each year we construct the risk set for the probability of continuing in the panel. A worker is assigned an outcome value of one if they remain in the PSID in a given year and assigned a zero in the year they attrite. A worker could then contribute at most 22 years of data and had to contribute at least one year of data (See Allison (1984) for a succinct discussion of discrete hazard models). The panel continuation hazard functions we estimated control for nonlinearity in age and education, length and other interview characteristics, race, poverty status, marital and family status, home ownership, location history, and time in the PSID.[16] Although time in the PSID may reflect duration dependence, because we do not formally control for latent heterogeneity the coefficient of time in the PSID does not have a single interpretation. Other research has found little evidence of latent heterogeneity in continuing in the PSID so that time in the panel should largely reflect duration dependence (Lillard and Panis 1994).

Panel Continuation Hazard Estimates. The likelihood of continuing in the PSID significantly increases at a decreasing rate as the participant ages in three of the eight specifications in Table 5. The likelihood of panel continuation increases with education, which is more apparent in the larger unbalanced panel models that control for completed participation. Location and socio-economic status have no estimated impact on whether a prime-aged man continues participating in the PSID. In general, the most important factors in terms of significant coefficients that are robust across models are interview

[16] To have time varying covariates in the hazard functions requires lagging regressors one year because information is not available in the year of attrition. Time varying covariates are troublesome because we lose the first year of data, and the bulk of persons who attrite contribute only one person year. To learn the consequences of using time-varying regressors for labor supply estimation we also estimated panel continuation hazards with only time invariant regressors, one for 1968–1989 and the other for 1969–1989. A Wald test indicates that the two sets of results are the same, which gives us confidence that our study of the consequences of panel attrition based on the panel continuation hazards with time varying regressors in Table 5 are not biased against finding non-random attrition.

characteristics, who was interviewed, and how long the respondent was already in the panel. For example, adding the number of waves as a regressor reduces the negative of the log-likelihood by 50 percent. If we draw on the results of Lillard and Panis (1994) who conclude that attrition from the PSID is adequately explained by measured covariates with no significant room remaining for latent individual heterogeneity, we can interpret the coefficient of Waves in Table 5 as indicating substantial duration dependence in continuing to participate in the PSID.

4.3 Two-Step Labor Supply Results

Tables 6 and 7 display two-step labor supply equation estimates corrected for the expected likelihood of continuing in the PSID. Because the presence of waves as a regressor in the continuation hazards may introduce endogeneity into the inverse Mills' ratio correction term in the labor supply equation we estimate the second step both with and without the waves regressor in Table 6. In Table 7 we restrict our attention to the probit and logit selection terms with waves as a regressor while letting the definition of wealth and wage instrument set differ. All models presented in Table 6 have been estimated with average hourly earnings as both the regressor and instrument set member with wealth the sum of liquid and illiquid assets. We emphasize that the key to understanding the econometric consequences of how (whether) one allows for possible non-random attrition from the PSID is not only whether the coefficient of the additional regressor capturing the probability of continuing to participate in the PSID is significant but also whether the economic coefficients of interest, particularly the estimated wage elasticities, change.

With exception of the labor supply results based on the discrete logit continuation hazards including waves estimated on the balanced panel, none of the selectivity terms is significantly different from zero in Tables 6 and 7. The estimated wage and asset coefficients from the balanced sample are sensitive to the choice of selectivity correction, asset measure, and wage instrument but tests of the differences across models are again going to be of low power because of the relatively small size of the balanced panel. Unlike the balanced panel results the larger unbalanced panel results are notably similar across selectivity terms, wage measures, and definitions of wealth. When wealth is measured as liquid assets alone and the wage instrument is average earnings per usual hours worked there is a relative loss of efficiency. It seems best to use the more comprehensive wealth measure including illiquid assets along with average hourly earnings in the instrument set. We note that the overidentifying restrictions are rejected for all the two-step labor supply models estimated with the unbalanced panel in Tables 6 and 7.

Table 8 displays results that further examine the relative importance of whether the researcher conditions for possible non-random attrition by using the subsample of years 1976–1989 when the preferred reported hourly wage rate measure is available in the PSID. The first four columns of Table 8 have no selectivity correction terms and the last four columns contain selectivity correction terms based on the discrete probit hazards with duration dependence presented in Table 5. Note that the columns of Table 8 differ by wage measure in the regression and instrument set. Finally, we examined the importance of latent heterogeneity to labor supply estimates by estimating labor supply under the null hypothesis of worker homogeneity (common intercepts) in the fourth and last columns of Table 8.

Comparing the columns labeled (2) and (3) to the columns labeled (1) in Table 8 illustrates the downward division bias inherent in labor supply functions estimated with the wage measured as average hourly earnings. Using average hourly earnings instead of the more accurate reported hourly wage reduces the estimated wage elasticites by 60–70 percent. The overidentifying restrictions are not rejected in labor supply functions permitting heterogeneity and using the reported hourly wage, but are rejected in all other cases in Table 8. The most important result in Table 8 is that the only time where the selectivity correction is statistically significant is in the specifications that improperly ignore latent labor supply heterogeneity. We conclude from our results using popular parametric specifications of the panel continuation process that any econometric bias that might result from ignoring attrition when estimating prime-aged males' labor supply with the PSID will be avoided by estimating a fixed-effects labor supply specification. The likelihood of attriting, while possibly endogenous, is largely person-specific and time-invariant so that fixed effects labor supply models obviate the need for two-step estimation with a first-step equation for the likelihood of continuing in the PSID.

5. Conclusion

We have examined the consequences of possible non-random panel attrition in a life cycle consistent model of labor supply permitting intertemporally progressive taxation of wage and interest incomes and latent worker specific heterogeneity. We examined Wald tests of whether the labor supply behavior of attriters is the same as non-attriters and a variable addition test involving two-step labor supply models that declare ex ante a discrete hazard function for panel continuation and then examine whether the labor supply coefficients of interest are affected significantly by the adding the likelihood of continuing in the panel. Our main conclusion is that alternative econometric specification decisions such as instrument set choice and wage regressor definition matter more than how (whether) one allows for the possible non-random attrition when estimating labor supply of prime-aged men with the PSID. Using a fixed effects labor supply equation conditions out any bias from possible non-random attrition.

References

Allison, Paul D. (1984). *Event History Analysis: Regression for Longitudinal Event Data*, Sage University Paper Series Number 46, Thousand Oaks, CA: Sage Publications, Inc.

Altonji, Joseph G. (1986). "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy* 94:3, S176–S215.

Becketti, Sean, Gould, William, Lillard, Lee and Finis Welch (1988). "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation," *Journal of Labor Economics* 6, 472–492.

Blomquist, N. Soren (1985). "Labor Supply in a Two-Period Model: The Effect of a Nonlinear Progressive Income Tax," *Review of Economic Studies* 70, 515–524.

Conway, Karen Smith and Thomas J. Kniesner (1994). "Estimating Labor Supply with Panel Data," *Economics Letters* 44, 27–33.

Davidson, Russell and James MacKinnon (1993). *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

Diggle, Peter J. and M.G. Kenward (1994). "Informative Drop-out in Longitudinal Data Analysis," *Applied Statistics* 43, 49–93.

Engen, Eric M., William Gale, and John K. Scholz (1994). "Do Saving Incentives Work?," *Brookings Papers on Economic Activity* 1, 85–151.

Greene, William H. (1993). *Econometric Analysis, Second Edition*, New York: Macmillan Publishing Company.

Griliches Zvi and Jerry Hausman (1986). "Errors in Variables in Panel Data," *Journal of Econometrics* 31, 93–118.

Hausman, Jerry and David Wise (1979). "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica* 47, 455–473.

Hausman, Jerry (1981). "Labor Supply," in *How Taxes Affect Economic Behavior*, H. Aaron and J. Pechman (eds.), Washington, DC: Brookings Institution.

Heckman, James J. (1979). "Sample Selection Bias as a Specification Error," *Econometrica* 47, 153–161.

Heckman, James J. and Thomas E. MaCurdy (1986). "Labor Econometrics," in *Handbook of Econometrics*, Vol. III, Z. Griliches and M. Intriligator (eds.), Amsterdam: North-Holland.

Hill, Martha S. (1992). *The Panel Study of Income Dynamics: A User's Guide*, Newbury Park, CA: Sage Publications.

Hoem, Jan M. (1989). "The Issue of Weights in Panel Surveys of Individuals' Behavior," in *Panel Surveys*, D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh, (eds.), New York: John Wiley and Sons, 539–565.

Hsiao, Cheng (1986). *Analysis of Panel Data*, New York: Cambridge University Press.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen Zeldes (1994). "Expanding the Life-Cycle Model: Precautionary Saving and Public Policy," *American Economic Review, Papers and Proceedings* 84:2, 174–179.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen Zeldes (1995). "Precautionary Saving and Social Insurance," *Journal of Political Economy*, 103:2, 360–399.

Johnson, William R. and Jonathan Skinner (1986). "Labor Supply and Marital Separation," *American Economic Review*, 76:3, 455–469.

Keane, Michael and David Runkle (1992). "On the Estimation of Panel-Data Models with Serial Correlation when the Instruments are not Strictly Exogenous," *Journal of Business and Economic Statistics* 10, 1–9.

Kniesner, Thomas J. and Qi Li (1996). "Semiparametric Panel Data Models with Heterogeneous Dynamic Adjustment: Theoretical Considerations and an Application to Labor Supply," Working Paper, Department of Economics, Indiana University, Bloomington, May.

Lillard, Lee A. (1989). "Sample Dynamics: Some Behavioral Issues," in *Panel Surveys*, D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh, (eds.), New York: John Wiley and Sons, 497–511.

Lillard, Lee A. and Constantijn W. A. Panis (1994). "Panel Attrition from the PSID: Household Income, Marital Status, and Mortality," Working Paper, The Rand Corporation, June.

Lo, Andrew and Whitney Newey (1985). “A Large Sample Chow Test for the Linear Simultaneous Equation,” *Economics Letters* 18, 351–353.

MaCurdy, Thomas E. (1983). “A Simple Scheme for Estimating An Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty,” *International Economic Review* 24, 265–289.

MaCurdy, Thomas, Green, David, and Harry Paarsch (1990). “Assessing Empirical Approaches for Analyzing Taxes and Labor Supply,” *Journal of Human Resources* 25, 415–490.

Maeshiro, Asatoshi and Shapoor Vali (1988). “Pitfalls in the Estimation of a Differenced Model,” *Journal of Business and Economic Statistics* 6, 511–515.

Manski, Charles (1989). “Anatomy of the Selection Problem,” *Journal of Human Resources* 24, 343–360.

Manski, Charles (1993). “The Selection Problem in Econometrics and Statistics,” in *Handbook of Statistics 11: Econometrics*, G.S. Maddala and C.R. Rao, (eds.), Amsterdam: North-Holland.

Newey, Whitney K. and Kenneth D. West (1987a). “Hypothesis Testing with Efficient Method of Moments Estimation,” *International Economic Review* 28, 777–787.

Newey, Whitney K. and Kenneth D. West (1987b). “A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica* 55, 703–708.

Robins, Philip K. and Richard W. West (1986). “Sample Attrition and Labor Supply Response in Experimental Panel Data,” *Journal Business and Economic Statistics* 4, 329–338.

Runkle, David E. (1991). “Liquidity Constraints and the Permanent Income Hypothesis: Evidence from Panel Data,” *Journal of Monetary Economics* 97, 73–98.

Verbeek, Marno and Theo Nijman (1992). “Testing for Selectivity Bias in Panel Data Models,” *International Economic Review* 33, 681–703.

Zabel, Jeffrey E. (1994). “An Analysis of Attrition in the PSID and SIPP with an Application to a Model of Labor Market Behavior,” presented at the BLS Conference on the Consequences of Attrition in Longitudinal Data, Washington, DC, February.

Ziliak, James P. (1994). “Does the Choice of Consumption Matter? An Application to the Permanent Income Hypothesis,” Working Paper 163, University of Oregon, Department of Economics.

Ziliak, James P. and Thomas J. Kniesner (1995). “Estimating Life-Cycle Labor Supply Tax Effects,” Discussion Paper No. 9589, CentER for Economic Research, Tilburg University, September.

Table 2: Summary Statistics for the Balanced and Unbalanced Samples for 1968-1989†

	Balanced SRC/SEO Sample††				Unbalanced SRC/SEO Sample††			
	Unweighted		Weighted		Unweighted		Weighted	
Variable	Non-Attriters	Attriters	Non-Attriters	Attriters	Non-Attriters	Attriters	Non-Attriters	Attriters
Anhh	2196.02	2171.97	2326.58	1882.83	2190.05	2181.72	2290.22	1914.01
	(420.58)	(526.78)	(460.49)	(616.55)	(478.08)	(553.49)	(529.54)	(647.81)
Riwag1	14.97	11.96	17.21	11.64	14.32	11.74	16.34	11.58
	(9.13)	(6.25)	(10.92)	(8.21)	(7.86)	(6.84)	(9.55)	(10.25)
Riatwg1	10.73	9.32	12.15	8.90	10.38	9.02	11.64	8.65
	(5.15)	(4.17)	(6.13)	(5.35)	(4.80)	(4.27)	(5.68)	(6.01)
Mtr	0.25	0.19	0.28	0.18	0.24	0.20	0.27	0.19
	(0.09)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)
Lasset	11.03	1.03	14.21	1.26	11.40	2.80	15.17	3.42
($1000's)	(45.66)	(5.80)	(55.62)	(8.21)	(42.43)	(16.69)	(55.87)	(25.85)
Equity	44.71	17.26	53.56	18.60	43.10	22.61	53.20	25.84
($1000's)	(50.90)	(46.90)	(59.05)	(68.57)	(53.02)	(41.18)	(67.99)	(59.52)
Age	43.50	38.08	45.65	33.01	43.48	39.03	45.00	34.77
	(8.41)	(7.28)	(9.30)	(9.14)	(8.15)	(6.99)	(9.25)	(8.77)
Kids	1.72	2.21	1.67	1.56	1.71	2.19	1.61	1.58
	(1.62)	(1.97)	(1.59)	(1.89)	(1.68)	(1.92)	(1.67)	(2.03)
Educ	11.08	10.76	12.36	10.05	11.11	10.39	12.10	9.61
	(4.29)	(3.50)	(4.65)	(4.25)	(3.89)	(3.73)	(4.52)	(4.84)
Race	0.82	0.62	0.92	0.72	0.71	0.55	0.92	0.67
	(0.38)	(0.49)	(0.24)	(0.49)	(0.45)	(0.50)	(0.38)	(0.57)
Home	0.81	0.56	0.87	0.53	0.79	0.58	0.86	0.60
	(0.39)	(0.50)	(0.40)	(0.58)	(0.41)	(0.49)	(0.41)	(0.57)
Married	0.92	0.83	0.98	0.75	0.91	0.86	0.96	0.78
	(0.27)	(0.37)	(0.23)	(0.39)	(0.28)	(0.35)	(0.29)	(0.38)
Waves	22.00	9.08	22.00	7.99	21.06	11.49	21.81	10.53
	(0.00)	(5.79)	(0.00)	(7.20)	(1.06)	(5.61)	(1.19)	(6.97)
Obs.	1958	711	1958	711	7100	1867	7100	1867

† Sample means are reported in the first row and standard deviations are in parentheses. The balanced sample deletes the entire time-series of persons if any missing values are encountered while in the sample; the unbalanced sample deletes only the person-year when missing values are encountered.
†† SRC=Survey Research Center; SEO=Survey of Economic Opportunity.

Table 5: Discrete Hazard Models for Probability of Panel Continuation†

	Balanced Panel Models				Unbalanced Panel Models			
Variable	Probit		Logit		Probit		Logit	
Constant	-3.973*	-0.044	-9.945**	0.654	-1.664	-0.212	-5.513**	-0.833
	(1.859)	(1.176)	(3.674)	(2.352)	(1.066)	(0.725)	(1.999)	(1.546)
Age	0.217**	0.001	0.508**	-0.036	0.053	0.046	0.201*	0.093
	(0.078)	(0.052)	(0.155)	(0.107)	(0.045)	(0.031)	(0.086)	(0.067)
Age^2	-0.003**	0.000	-0.008**	0.001	-0.001	-0.001	-0.003**	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Educ	0.029	0.017	0.087	0.001	0.095*	0.033	0.210**	0.067
	(0.098)	(0.055)	(0.192)	(0.108)	(0.041)	(0.030)	(0.080)	(0.069)
Aged	-0.002	-0.000	-0.004	0.000	-0.003**	-0.001	-0.006**	-0.002
	(0.002)	(0.001)	(0.005)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
Race	-0.173	0.323*	-0.186	0.587*	-0.191	0.119	-0.286	0.247
	(0.181)	(0.130)	(0.335)	(0.248)	(0.108)	(0.073)	(0.204)	(0.158)
Seo	0.149	0.153	0.389	0.348	0.001	0.042	0.023	0.116
	(0.184)	(0.126)	(0.341)	(0.244)	(0.107)	(0.073)	(0.204)	(0.157)
Kids	-0.048	0.011	-0.065	0.007	0.025	0.012	0.046	0.027
	(0.043)	(0.026)	(0.084)	(0.051)	(0.027)	(0.018)	(0.050)	(0.038)
Phone	-1.252**	0.862**	-2.552**	1.752**	-0.700**	0.532**	-1.479**	1.169**
	(0.223)	(0.112)	(0.448)	(0.239)	(0.140)	(0.078)	(0.276)	(0.169)
Head	0.197	0.414**	0.432	0.833**	0.196	0.329**	0.372	0.726**
	(0.202)	(0.132)	(0.387)	(0.253)	(0.111)	(0.077)	(0.218)	(0.163)
Married	0.336	0.174	0.741	0.350	0.206	0.223*	0.486*	0.470*
	(0.214)	(0.146)	(0.411)	(0.274)	(0.125)	(0.089)	(0.239)	(0.186)
Mover	0.056	0.468**	-0.166	0.866**	0.053	0.144	-0.008	0.306
	(0.227)	(0.154)	(0.411)	(0.301)	(0.128)	(0.085)	(0.239)	(0.179)
Wmove	-0.148	-0.234*	-0.269	-0.437**	-0.075	-0.113	-0.153	-0.243
	(0.154)	(0.097)	(0.285)	(0.186)	(0.099)	(0.063)	(0.185)	(0.134)
Intlg	-0.001	-0.001	-0.002	-0.002	0.005*	-0.000	0.009*	-0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.004)	(0.003)
Home	0.079	0.288**	0.098	0.573**	-0.038	0.285**	-0.108	0.619**
	(0.177)	(0.107)	(0.331)	(0.207)	(0.110)	(0.068)	(0.208)	(0.148)
Calls	-0.379**	-0.362**	-0.595**	-0.720**	-0.341**	-0.276**	-0.592**	-0.609**
	(0.105)	(0.065)	(0.198)	(0.125)	(0.062)	(0.040)	(0.119)	(0.085)
Neast	-0.152	0.117	-0.271	0.186	0.022	-0.014	0.029	-0.018
	(0.217)	(0.137)	(0.404)	(0.264)	(0.137)	(0.092)	(0.259)	(0.197)
Ncent	0.039	-0.153	0.045	-0.328	-0.010	0.004	0.040	0.049
	(0.213)	(0.134)	(0.392)	(0.257)	(0.134)	(0.086)	(0.257)	(0.186)
South	-0.542*	0.116	-0.858*	0.134	-0.211	0.059	-0.332	0.154
	(0.213)	(0.128)	(0.394)	(0.246)	(0.121)	(0.081)	(0.233)	(0.175)
Waves	0.353**		0.703**		0.265**		0.533**	
	(0.026)		(0.056)		(0.012)		(0.025)	
LL	-216.859	-552.550	-210.195	-552.291	-574.976	-1253.85	-570.448	-1253.52
Obs.	2580.0	2580.0	2580.0	2580.0	8659.0	8659.0	8659.0	8659.0

† Standard errors are given in parentheses. The balanced sample deletes the entire time-series of persons if any missing values are encountered while in the sample; the unbalanced sample deletes only the person-year when missing values are encountered.

* denotes significance at the 5 percent level; ** denotes significance at the 1 percent level; LL is the value of the log-likelihood function at the maximum

Table 1: Definitions of Variables Used in Estimation†

Anhh:	Annual hours of work
Riwag1:	Real gross hourly wage defined as (annual earnings)/(annual hours), known as the imputed wage
Riwag2:	Real gross hourly wage defined as (annual earnings)/(weeks worked* usual hours per week), known as the weeks-worked wage
Rrwag:	Real gross hourly wage defined as the reported hourly or salary wage rate, known as the reported wage
Riatwg1:	Real after-tax wage rate defined as riwag1*(1-mtr)
Riatwg2:	Real after-tax wage rate defined as riwag2*(1-mtr)
Rratwg:	Real after-tax wage rate defined as rrwag*(1-mtr)
Mtr:	marginal tax rate defined as a cubic polynomial in taxable income
Lasset:	Real liquid assets defined as the ratio of rent,interest,dividend income over nominal interest rate
Equity:	Real home equity defined as house value less principal remaining
Assets:	Combined values of Lasset and Equity
Age:	Age of the male head of household
Kids:	The number of children residing in the household
Educ:	The number of years of schooling for the head
Aged:	Interaction of age*educ
Race:	A dummy variable equal to 1 if the head is white
Home:	A dummy variable equal to 1 if the head owns his home
Married:	A dummy variable equal to 1 if the head is married
Seo:	A dummy variable equal to 1 if the head is part of the SEO subsample
Head:	A dummy variable equal to 1 if the head was the respondant for the interview
Phone:	A dummy variable equal to 1 if the interview was conducted by telephone
Intlg:	The length of the interview
Mover:	A dummy variable equal to 1 if the head moved since last interview
Wmove:	A dummy variable equal to 1 if the head is thinking of a move soon
Calls:	The natural log of the number of phone calls required to reach the respondant
Neast:	A dummy variable equal to 1 if the head resides in the Northeast
Ncent:	A dummy variable equal to 1 if the head resides in the NorthCentral
South:	A dummy variable equal to 1 if the head resides in the South
West:	A dummy variable equal to 1 if the head resides in the West, Alaska, or Hawaii
Waves:	The number of waves the panel participant was in the sample
IMR:	The inverse of Mill’s Ratio

† All wage and wealth variables are deflated by the 1987 Personal Consumption Expenditure Deflator

Table 3: Wald Tests Comparing Attriters and Non-Attriters in a Life-Cycle Consistent Labor Supply Model for the Years 1968-1989†

	Balanced SRC/SEO Sample††				Unbalanced SRC/SEO Sample††			
Variable	Non-Attriters (1)	Attriters (1)	Non-Attriters (2)	Attriters (2)	Non-Attriters (1)	Attriters (1)	Non-Attriters (2)	Attriters (2)
Riatwg1 (α)	20.542 (21.969)	-100.034** (35.197)	19.056 (26.579)	-58.062 (12.4E2)	22.634* (9.632)	-48.384* (23.081)	19.898* (9.954)	-742.003** (29.296)
Lagged Assets (δ)	0.416 (0.442)	-3.825 (2.698)	-0.309 (1.224)	5.315 (8.9E2)	0.653* (0.274)	-1.062 (0.792)	0.489 (0.310)	20.817** (1.066)
Current Assets (ϕ)	0.167 (1.737)	3.113 (4.932)	-1.493 (2.666)	-22.966 (4.9E2)	1.483 (0.896)	-8.832** (2.997)	-1.097 (1.014)	-252.657** (7.781)
Uncomp. Wage	0.103 (0.110)	-0.492** (0.173)	0.096 (0.133)	-0.286 (6.102)	0.109* (0.047)	-0.215* (0.102)	0.097* (0.048)	-3.297** (0.130)
Comp. Wage	0.101 (0.112)	-0.524** (0.180)	0.112 (0.136)	-0.051 (7.928)	0.094* (0.048)	-0.131 (0.106)	0.108* (0.049)	-0.889** (0.149)
J-statistic [dof, p]	21.470 [33,.939]	11.676 [30,.998]	13.317 [33,0.999]	0.004 [30,.999]	62.129 [33,.002]	15.734 [32,.993]	61.121 [33,.002]	--.----
Wald test [dof, p]		9.799 [7,0.200]		0.005 [7,.999]		25.990 [7,.001]		1095.32 [7,0.0]
Obs.	1691.0	369.0	1691.0	369.0	6086.0	1186.0	6086.0	1186.0

† Standard errors are given in parentheses. The balanced sample deletes the entire time-series of persons if any missing values are encountered while in the sample; the unbalanced sample deletes only the person-year when missing values are encountered.
†† The imputed wage, Riatwg1, is used both as a regressor and as an instrument. Results in columns labelled (1) are based on Assets=Lasset+Equity; results in column (2) are based on Assets=Lasset.
* denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table 4: Wald Tests Comparing Attriters and Non-Attriters in a Life-Cycle Consistent Labor Supply Model for the Years 1968-1989†

	Balanced SRC/SEO Sample††				Unbalanced SRC/SEO Sample††			
Variable	Non-Attriters (1)	Attriters (1)	Non-Attriters (2)	Attriters (2)	Non-Attriters (1)	Attriters (1)	Non-Attriters (2)	Attriters (2)
Riatwg1 (α)	14.587 (130.688)	-159.577 (----.----)	-129.852** (17.065)	-30.006 (3.9E2)	18.214 (9.656)	-42.433 (23.194)	14.914 (9.938)	-51.732* (21.641)
Lagged Assets (δ)	-0.782 (7.270)	-10.328 (18.567)	4.515** (0.438)	-5.829 (1.3E2)	0.423 (0.275)	-1.439 (0.782)	0.432 (0.306)	0.710 (0.902)
Current Assets (ϕ)	-0.556 (10.584)	10.304 (28.610)	27.898** (2.551)	-57.042 (6.3E2)	0.872 (0.935)	-8.021** (2.941)	-1.095 (1.014)	-7.048* (2.815)
Uncomp. Wage	0.073 (0.655)	-0.785 (--.----)	-0.651** (0.088)	-0.148 (1.928)	0.088 (0.047)	-0.189 (0.103)	0.072 (0.048)	-0.229* (0.096)
Comp. Wage	0.079 (0.665)	-0.890 (--.----)	-0.956** (0.089)	0.435 (2.034)	0.079 (0.048)	-0.112 (0.107)	0.084 (0.049)	-0.163 (0.099)
J-statistic [dof, p]	0.951 [33,.999]	0.462 [30,.999]	--.----	0.188 [30,.999]	60.799 [33,.002]	25.754 [32,.774]	57.545 [33,.005]	10.841 [32,.999]
Wald test [dof, p]		--.----		1.716 [7,.974]		21.541 [7,.003]		19.036 [7,.008]
Obs.	1691.0	369.0	1691.0	369.0	6086.0	1186.0	6086.0	1186.0

† Standard errors are given in parentheses. The balanced sample deletes the entire time-series of persons if any missing values are encountered while in the sample; the unbalanced sample deletes only the person-year when missing values are encountered.
†† The imputed wage, Riatwg1, is used as a regressor and the weeks-worked wage, Riatwg2, is used as an instrument. Results in columns labelled (1) are based on Assets=Lasset+Equity; results in column (2) are based on Assets=Lasset.
* denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table 6: Two-Step Attrition Tests in a Life-Cycle Consistent Labor Supply Model for the Years 1968-1989†

	Balanced SRC/SEO Sample††				Unbalanced SRC/SEO Sample††			
Variable	Probit w/ Waves	Probit w/o Waves	Logit w/ Waves	Logit w/o Waves	Probit w/ Waves	Probit w/o Waves	Logit w/ Waves	Logit w/o Waves
Riatwg1	29.280	-71.057	-474.509**	13.584	22.356*	23.291*	21.329*	23.732*
(α)	(20.309)	(3.1E2)	(34.518)	(28.302)	(9.588)	(9.702)	(9.632)	(9.746)
Lagged	0.175	1.678	10.839**	-0.044	0.634*	0.629*	0.636*	0.607*
Assets (δ)	(0.446)	(5.821)	(0.759)	(0.769)	(0.271)	(0.274)	(0.268)	(0.274)
Current	-0.164	9.701	22.344**	-0.487	1.445	1.396	1.419	1.343
Assets (ϕ)	(1.654)	(26.228)	(1.927)	(2.147)	(0.899)	(0.905)	(0.891)	(0.908)
Uncomp.	0.147	-0.356	-2.379**	0.068	0.109*	0.113*	0.104*	0.115*
Wage	(0.102)	(1.566)	(0.173)	(0.142)	(0.047)	(0.047)	(0.047)	(0.047)
Comp.	0.149	-0.462	-2.623**	0.073	0.093*	0.098*	0.089	0.101*
Wage	(0.103)	(1.569)	(0.174)	(0.144)	(0.048)	(0.048)	(0.048)	(0.048)
IMR	-15.319	18.099	-31.8E2**	-101.456	-1.208	10.039	57.483	-31.426
	(12.472)	(4.3E2)	(5.8E2)	(104.789)	(3.902)	(14.574)	(121.629)	(31.149)
J-statistic	24.501	0.002	--.-----	14.020	62.293	61.362	63.504	60.702
[dof, p]	[32,.939]	[32,.999]		[32,.997]	[32,.001]	[32,.001]	[32,.001]	[32,0.002]
Obs.	1691.0	1691.0	1691.0	1691.0	6086.0	6086.0	6086.0	6086.0

† Standard errors are given in parentheses. The balanced sample deletes the entire time-series of persons if any missing values are encountered while in the sample; the unbalanced sample deletes only the person-year when missing values are encountered.

†† The imputed wage, Riatwg1, is used both as a regressor and as an instrument. Assets=Lasset+Equity.

* denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table 7: Two-Step Attrition Tests in a Life-Cycle Consistent Labor Supply Model for the Years 1968-1989†

	Balanced SRC/SEO Sample††				Unbalanced SRC/SEO Sample††			
Variable	Probit w/ Waves (1)	Logit w/ Waves (1)	Probit w/ Waves (2)	Logit w/ Waves (2)	Probit w/ Waves (1)	Logit w/ Waves (1)	Probit w/ Waves (2)	Logit w/ Waves (2)
Riatwg1	31.074	16.061	26.147	8.049	19.133	17.913	17.477	16.620
(α)	(7.6E2)	(46.450)	(124.1E2)	(84.658)	(10.018)	(10.017)	(9.603)	(9.575)
Lagged	0.221	0.202	0.519	0.716	0.465	0.482	0.410	0.413
Assets (δ)	(14.414)	(2.179)	(2.9E2)	(3.088)	(0.325)	(0.308)	(0.279)	(0.274)
Current	-1.045	-1.056	0.882	0.975	-1.189	-1.118	0.842	0.823
Assets (ϕ)	(83.093)	(5.287)	(8.8E2)	(5.138)	(1.045)	(1.008)	(0.945)	(0.929)
Uncomp.	0.155	0.081	0.131	0.040	0.093	0.087	0.085	0.081
Wage	(3.815)	(0.233)	(62.204)	(0.424)	(0.049)	(0.049)	(0.047)	(0.046)
Comp.	0.167	0.092	0.121	0.030	0.105*	0.099*	0.076	0.072
Wage	(3.922)	(0.240)	(62.939)	(0.428)	(0.050)	(0.050)	(0.048)	(0.047)
IMR	-22.642	4.5E2	-13.271	56.1E2	-2.841	92.238	-2.016	70.619
	(4.9E2)	(13.6E2)	(51.9E2)	(162.9E2)	(4.092)	(123.971)	(3.921)	(121.630)
J-statistic	0.014	4.160	0.000	2.043	60.217	61.896	60.885	62.489
[dof, p]	[32,.999]	[32,.999]	[32,1.00]	[32,.999]	[32,.002]	[32,.001]	[32,.002]	[32,0.001]
Obs.	1691.0	1691.0	1691.0	1691.0	6086.0	6086.0	6086.0	6086.0

† Standard errors are given in parentheses. The balanced sample deletes the entire time-series of persons if any missing values are encountered while in the sample; the unbalanced sample deletes only the person-year when missing values are encountered.

†† Results in columns labelled (1) are based on Assets=Lasset with the imputed wage, Riatwg1, as both the regressor and the instrument; results in column (2) are based on Assets=Lasset+Equity with the imputed wage, Riatwg1, as the regressor and the weeks-worked wage, Riatwg2, as the instrument.

* denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.

Table 8: Two-Step Attrition Tests in a Life-Cycle Consistent Labor Supply Model for the Unbalanced SRC/SEO Sample Over theYears 1976-1989†

Variable††	(1)	(2)	(3)	(4)	Probit w/ Waves (1)	Probit w/ Waves (2)	Probit w/ Waves (3)	Probit w/ Waves (4)
Net Wage	51.978**	21.477*	15.915	62.114**	55.612**	21.369*	17.477	29.209**
(α)	(16.344)	(9.132)	(9.059)	(3.866)	(17.611)	(9.317)	(9.603)	(3.415)
Lagged	0.313	0.452	0.196	1.678**	0.332	0.453	0.410	1.004**
Assets (δ)	(0.285)	(0.301)	(0.305)	(0.533)	(0.291)	(0.302)	(0.279)	(0.408)
Current	0.715	0.973	0.265	-4.062**	0.879	1.009	0.842	-1.334*
Assets (ϕ)	(1.036)	(1.061)	(1.073)	(0.711)	(1.081)	(1.082)	(0.945)	(0.557)
Uncomp.	0.238**	0.106*	0.079	0.307**	0.255**	0.106*	0.085	0.144**
Wage	(0.075)	(0.045)	(0.045)	(0.019)	(0.080)	(0.046)	(0.047)	(0.017)
Comp.	0.231**	0.096*	0.072	0.350**	0.246**	0.095*	0.076	0.156**
Wage	(0.076)	(0.047)	(0.048)	(0.021)	(0.081)	(0.047)	(0.048)	(0.018)
IMR					5.852	0.914	-2.016	313.769**
					(9.383)	(8.398)	(3.921)	(13.855)
J-statistic	37.059	50.326	49.626	364.128	36.359	50.656	60.885	177.028
[dof, p]	[25,.057]	[25,.002]	[25,.002]	[26,.000]	[24,.051]	[24,.001]	[32,.002]	[25,.000]
Obs.	3489.0	3489.0	3489.0	3827.0	3489.0	3489.0	3827.0	3827.0

† Standard errors are given in parentheses. The balanced sample deletes the entire time-series of persons if any missing values are encountered while in the sample; the unbalanced sample deletes only the person-year when missing values are encountered.

†† All results are based on Assets=Lasset+Equity. Results in columns labelled (1) are based on the reported wage, Rratwg; results in column (2) are based on the imputed wage, Riatwg1, as both the regressor and instrument; results in column (3) are based on the imputed wage, Riatwg1, as the regressor and the weeks-worked wage, Riatwg2, as the instrument; results in column (4) are based on the reported wage, Rratwg, for the model estimated in levels under the null hypothesis of no latent heterogeneity.

* denotes significance at the 5 percent level; ** denotes significance at the 1 percent level.